

# **User's Guide for the ECOSAR Class Program**

MS-Windows Version 0.99d

November 1998

*prepared by:*

William M. Meylan and Philip H. Howard

Syracuse Research Corporation  
Environmental Science Center  
6225 Running Ridge Road  
North Syracuse, NY 13210

*prepared for:*

J. Vincent Nabholz and Gordon Cash

Risk Assessment Division (7403)  
U.S. Environmental Protection Agency  
401 M St., SW  
Washington, DC 20460



## Table of Contents

	Page
1. INTRODUCTION .....	1
2. COMPUTER-SOFTWARE REQUIREMENTS .....	2
3. INSTALLING the ECOSAR Class Program .....	3
4. STARTING ECOSAR Class Program .....	5
5. DATA ENTRY and EDIT KEYS .....	7
5.1. Entering Data .....	7
5.1.1. SMILES Notation .....	7
5.1.2. Individual Data Entry Fields .....	7
5.2. Function Keys & Buttons .....	9
5.3. Importing Structures .....	11
6. RESULTS .....	13
6.1. Structure Window .....	14
6.2. Konemann Equation .....	15
6.3. Example SAR Equations .....	15
7. BATCH RUNS .....	16
7.1. Batch Output Formats .....	17
8. SPECIAL CLASS Calculations .....	18
9. BIBLIOGRAPHY .....	20
APPENDIX A - Selected SMILES Information .....	21
APPENDIX B - Description of User Input File .....	22
APPENDIX C - CAS Number Data Base .....	22
APPENDIX D - Estimation of Water Solubility .....	23
APPENDIX E - List of ECOSAR Chemical Classes .....	24

## 1. INTRODUCTION

The structure-activity relationships (SARs) presented in this program are used to predict the aquatic toxicity of chemicals based on their similarity of structure to chemicals for which the aquatic toxicity has been previously measured. Most SAR calculations in the ECOSAR Class Program are based upon the octanol/water partition coefficient ( $K_{ow}$ ). Various surfactant SAR calculations are based upon the average length of carbon chains or the number of ethoxylate units.

SARs have been used by the U.S. Environmental Protection Agency since 1981 to predict the aquatic toxicity of new industrial chemicals in the absence of test data. The acute toxicity of a chemical to fish (both fresh and saltwater), water fleas (daphnids), and green algae has been the focus of the development of SARs, although for some chemical classes SARs are available for other effects (e.g. chronic toxicity and bioconcentration factor) and organisms (e.g., earthworms). SARs are developed for chemical classes based on measured test data that have been submitted by industry or they are developed by other sources for chemicals with similar structures, e.g., phenols. Using the measured aquatic toxicity values and estimated  $K_{ow}$  values, regression equations can be developed for a class of chemicals. Toxicity values for new chemicals may then be calculated by inserting the estimated  $K_{ow}$  into the regression equation and correcting the resultant value for the molecular weight of the compound.

To date, over 150 SARs have been developed for more than 50 chemical classes. These chemical classes range from the very large, e.g., neutral organics, to the very small, e.g., aromatic diazoniums. Some chemical classes have only one SAR, such as acid chlorides, for which only a fish 96-hour  $LC_{50}$  has been developed. The class with the greatest number of SARs is the neutral organics, which has SARs ranging from acute and chronic SARs for fish to a 14-day  $LC_{50}$  for earthworms in artificial soil.

The ECOSAR Class Program is a computerized version of the ECOSAR analysis procedures as currently practiced by the Office of Pollution Prevention and Toxics (OPPT). It has been developed within the regulatory constraints of the Toxic Substances Control Act (TSCA). It is a pragmatic approach to SAR as opposed to a theoretical approach.

This ECOSAR program is designed for the expert user. You are expected to have some knowledge of environmental toxicology and organic chemistry. It is menu-driven and contains various help functions to assist you. You cannot change any of the equations or data stored within the program or accidentally erase any important information. The following pages show you

how to install, access, and use the ECOSAR program. If you have any questions or comments on the ECOSAR program, or find any errors, please contact:

ECOSAR Program  
Risk Assessment Division (7403)  
U.S. Environmental Protection Agency  
401 M St., SW  
Washington, DC 20460

## **2. COMPUTER-SOFTWARE REQUIREMENTS**

The ECOSAR Class Program is designed for use on the IBM and IBM-compatible series of personal computers running Microsoft Windows 3.1 and higher (including Windows 95 and Windows NT). Although a mouse or other pointing device is not required, it is highly recommended. The ECOSAR Class Program requires approximately 0.5 MB of hard disk space. Use of the supplemental SMILECAS Database (a database of 103,000 SMILES notations indexed by CAS number for program retrieval to ECOSAR Class Program) requires a hard drive and  $\approx 9.7$  MB of additional disk space.

The ECOSAR Class Program runs under Windows 95/98/NT; however, it is not currently designed to run as a multi-tasking program (e.g. running the ECOSAR Class Program batch-mode runs in the background while running another program in the foreground). Batch-modes should be run in the foreground until completion.

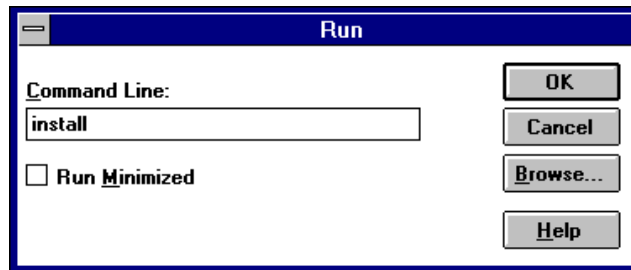
### 3. INSTALLING the ECOSAR Class Program

The ECOSAR Class Program diskette contains an installation program that can install ECOSAR Class Program and create a Windows Program Group with program icon. The installation program must be started while Microsoft Windows (3.1, 95 or NT) is running. To install, place the floppy diskette in the appropriate floppy drive. Then, (a) in Windows 3.1, select FILE, RUN from the Program Manager's menu, or (b) in Windows 95/98, press the Start button and select Run. Then:

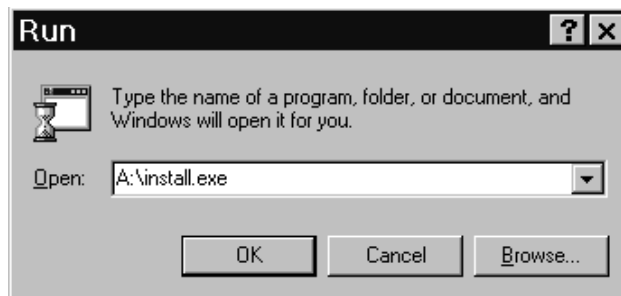
If the floppy is in the a: drive, enter **a:install**

If the floppy is in the b: drive, enter **b:install**

The FILE, RUN entry box (in Windows 3.1) may look similar to the following:



The Run entry box (in Windows 95/98) may look similar to the following:



The ECOSAR Class Program program does not actually require the installation program because installation can be handled manually; ECOSAR Class Program and its help file can be used as they exist on the floppy (that is, you can start ECOSAR Class Program directly from the

floppy if you want). However, the installation program automatically creates a hard-drive subdirectory, copies the necessary files to it, and creates a Windows program group. The ECOSAR Class Program group folder (named "Ecosar") contains a ECOSAR icon that starts the program.

The following files are installed during the installation process:

- ECOWIN.EXE: the necessary ECOSAR executable file
- ECOWHELP.HLP: a file containing extensive help information for SMILES notations, program execution, key & button usage, etc.

The following files are NOT installed during the installation process (and are not on the installation disk), but can be used by the ECOSAR Class Program. These files must be obtained separately:

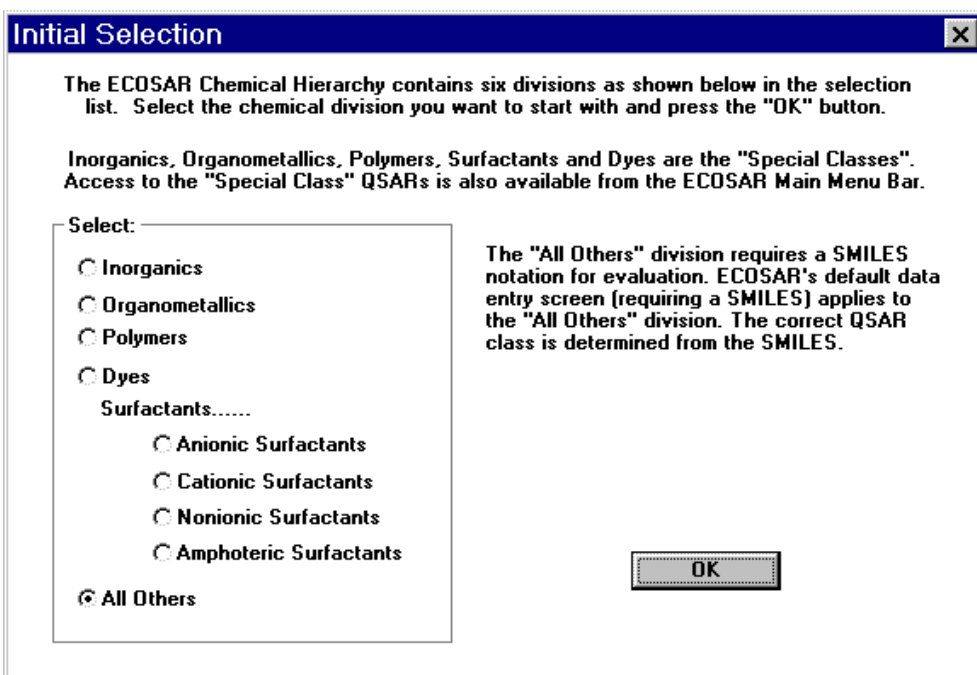
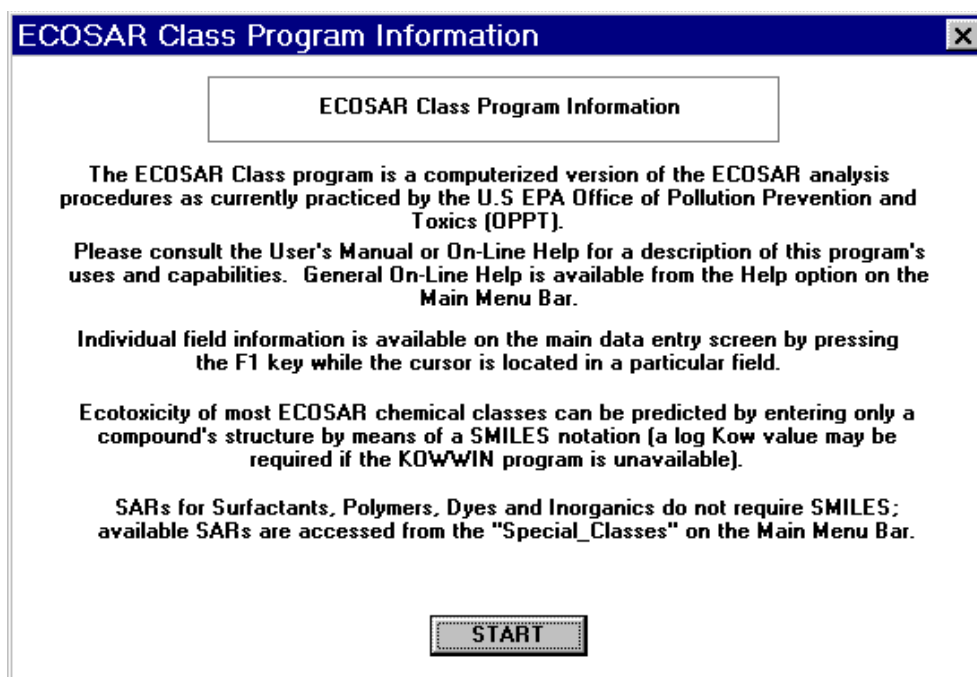
- SMILECAS.DB: a database of more than 103,000 SMILES notations indexed by CAS (Chemical Abstract Registry) number. By simply entering a CAS number in ECOSAR allows automatic retrieval of available SMILES. This database can also be used to run automated batches of CAS numbers.
- SMILECAS.IDX: index file for SMILECAS.DB
- KOWWIN.EXE: a Syracuse Research Corporation program that estimates log  $K_{ow}$  from SMILES. When this program (and its two library files listed below) are available in the same subdirectory as ECOSAR, it allows ECOSAR to automatically start KOWWIN and retrieve the KOWWIN estimate. The estimation methodology is described in a journal article (Meylan and Howard, 1995).
- Note:** the KOWWIN program located in the ECOSAR subdirectory must be closed while ECOSAR is running. Otherwise, ECOSAR can not use it. A duplicate copy of KOWWIN can be running from a different subdirectory however.
- CSDLL.DLL: a library file required by the KOWWIN program.
- QCBASED.DLL: a library file required by the KOWWIN program.

These files are available from Syracuse Research Corporation, Environmental Science Center, 6225 Running Ridge Road, North Syracuse, NY 13212-2510 (Dr. Phil Howard, 315-452-8417).

#### 4. STARTING the ECOSAR Class Program

The ECOSAR Class Program is started like any other Microsoft Windows program. The easiest way to start ECOSAR Class Program is to double-click the program icon installed in the ECOSAR program group during installation. For additional information on starting Windows programs, consult your Windows documentation.

The following Introductory Screens are displayed:



The “Initial Selection” screen lists the six divisions of the ECOSAR Chemical Hierarchy. The divisions are Inorganics, Organometallics, Polymers, Dyes, Surfactants (Anionic, Cationic, Nonionic, Amphoteric), and All Others. Inorganics, Organometallics, Polymers, Dyes and Surfactants are the “Special Classes” (see Section 8). The default selection is “All Others”. This division requires a SMILES notation for evaluation. Appendix E lists the Chemical Classes identified in the “All Others” division.

Program execution (“All Others” division) begins at the data entry screen; an example is illustrated in Figure 1.

**Ecosar Classes v0.99d**

File Edit Functions BatchMode ShowStructure Special\_Classes Help

Previous Get User Save User CAS Input Calculate

Enter SMILES:

Enter NAME:

CAS Number:

Chemical ID 1:

Chemical ID 2:

Chemical ID 3:

Measured Water Sol (mg/L):

Melting Point (deg C):

Log Kow:

Measured Log Kow:

**Figure 1.** Example Data Entry Screen

Note: the appearance of the screen may vary somewhat due to screen resolution (e.g. 640 X 480 vs. 800 X 600), user selection of MS-Windows attributes (e.g. colors, font size, etc.), etc. In addition, Figure 1 illustrates how the entry screen appears when using Windows 95. Appearance in Windows 3.1 varies slightly.



## 5. DATA ENTRY and EDIT KEYS

The information in section 5 applies to the main data entry screen shown in Figure 1. It concerns structure estimation using SMILES notation. Information concerning data entry for "Special Classes" (calculations not using SMILES) is presented in section 8.

### 5.1. Entering Data

#### 5.1.1. SMILES Notation

Calculations from the main data entry screen require the chemical structure of the compound as a SMILES notation. Users unfamiliar with SMILES notations can consult a descriptive journal article (Weininger, 1988) or the ECOSAR Class Program help file (accessed by selecting "Help" from the top menu). The following Internet web-site locations also contain extensive information about SMILES notations:

- (1) <http://www.daylight.com> (Daylight Information Services)
- (2) <http://esc.syrres.com> (Syracuse Research Corporation)

Three different methods can be used to enter the SMILES notation with chemical name:

- (1) direct entry by the user from the keyboard
- (2) entry from a previously created user file that is accessed by pressing the F4 key (or clicking the "Get User" button)
- (3) entry from a supplementary database that is accessed by pressing the F8 key (or clicking the "CAS Input" button) and entering the Chemical Abstract Service (CAS) Registry number of the compound.

The program can estimate only one chemical at a time. Separate data entry is required for each chemical, although batch mode runs are possible (see F5, F7 function keys below and see Section 7).

Estimation of the entered SMILES notation is started by pressing the PgDn key (or clicking the "Calculate" button) at any time during data entry.

#### 5.1.2. Individual Data Entry Fields

The following is a description of the individual data entry fields on the main data entry screen (pressing the F1 key where the edit cursor is located gives a brief description of that field):

- (1) **SMILES:** the SMILES notation of the structure to be estimated. A maximum of 360 characters are allowed. This field is required. Do not leave any blank spaces in front of a SMILES notation ... a SMILES is considered finished when a blank space is encountered.
- (2) **Name:** the name and/or description of the structure. This field is optional; not required. A maximum of 120 characters are allowed.
- (3) **CAS Number:** the CAS (Chemical Abstract Service Registry) Number. This field is optional; not required. When a SMILES is retrieved from the SMILECAS Database, the CAS is automatically inserted in this field.
- (4) **Chemical ID 1:** optional description / identity field; not required.
- (5) **Chemical ID 2:** optional description / identity field; not required.
- (6) **Chemical ID 3:** optional description / identity field; not required.
- (7) **Log Kow:** the log octanol-water partition coefficient. A value is required unless the KOWWIN Program (Syracuse Research Corporation) is present in the same subdirectory as the ECOSAR Class Program. When KOWWIN (and its two library files) are available in the same subdirectory as ECOSAR, it allows ECOSAR to automatically start KOWWIN and retrieve the KOWWIN estimate. Note: the KOWWIN program located in the ECOSAR subdirectory must be closed while ECOSAR is running. Otherwise, ECOSAR can not use it. A duplicate copy of KOWWIN can be running from a different subdirectory however.
- (8) **Measured Water Solubility:**  
the Measured Water Solubility in mg/L. This field is optional. It is NOT required! When left blank, a Water Solubility will be calculated from the log Kow value. Predicted toxicity values are compared to the Water Solubility .... if toxicity exceeds Water Solubility, the toxicity value is marked with an asterick (\*) to indicate 'No Effect at Saturation'. Water Solubility is not used to calculate ecotoxicity values. The estimation methodology is described in Appendix D.
- (9) **Melting Point:** the Melting Point (in deg C). This field is optional; not required. It is used to calculate Water Solubility when a measured Water Solubility is unavailable. It generally helps in estimating more accurate water solubilities, but is not required to estimate Water Solubility.
- (10) **Measured Log Kow:**  
the measured log Kow value, if available. This field is informational only. It is not used to calculate ecotoxicity values. The value in the Log K<sub>ow</sub> field is used to calculate ecotoxicity values.

## 5.2. Function Keys & Buttons

**F1:** Accesses a help message for the individual field where the blinking cursor is located. General Help is available from "Help" on the Menu Bar at the top of the screen. It is a standard Windows help system; to access a specific help topic, simply click on the topic (or keyword) that is highlighted in green where the mouse pointer changes to a hand.

Previous

**F2:** Pressing the F2 key or clicking the "Previous" button recalls the most recent SMILES and chemical name that was calculated or attempted to be calculated by the program. It can save a lot of time when making small changes to large SMILES and names. It is especially useful after a SMILES notation error occurs....the incorrect SMILES can be recalled and edited.

**F3:** Clears the currently displayed SMILES Notation, Chemical Name and other data. All entry fields are filled with blank spaces.

Get User

**F4:** Pressing the F4 key or clicking the "Get User" button displays a file selection dialog box that allows the user to open a file of previously saved SMILES notations and chemical names. The default name of the file is SMILES.INP; this is for compatibility with similar programs. The file selection box looks for files with the extension ".INP", so it is best to name files with this extension when creating them with the F6 key ("Save User"). A "Get User" file can contain up to 1500 SMILES and names and the user can select any single SMILES and name for input. The

SMILES.INP file can be created one chemical at a time by using the F6 key as described below. Also, "Get User" option is only usable after a file has been created with the F6 key feature!! An example screen is shown to the right. Selection is made by highlighting the desired line and clicking to "OK" button or by double-clicking the desired line. See Appendix B for the correct file format required!

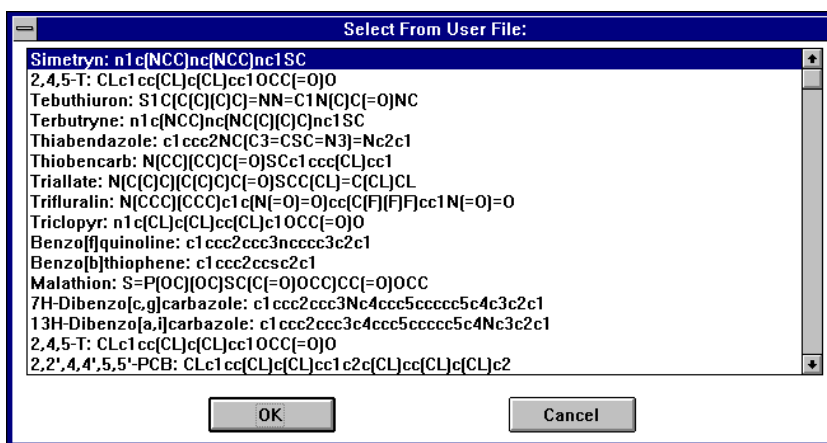


Figure 2. Example User Input File Selection

**F5:** Pressing the F5 key (or clicking the "BatchMode" option on the main menu and selecting "Batch File Input Using SMILES Strings") brings up the selection box shown. The F5 key is used for batch entry of SMILES strings from ascii text files. The text files MUST be in either of two formats. (1) String Format or (2) EcowinFormat.

String Format must have the SMILES string at the beginning of each line in the file; it can then be followed by a space(s) and then the name or other ID. The SMILES is considered terminated at the first space. An example String Format is as follows:

```
CCCCO Butanol
c1ccccc1 Benzene
Fc1ccccc1 Fluorobenzene
CC(=O)C Acetone
```

EcowinFormat is the same format used by the "Get User" and "Save User" button features. Therefore, the "SMILES.INP" file can be used directly to run batch file outputs. In this format, the name comes first (maximum of 60 characters) followed by a colon and one space, and then the SMILES notation. An example EcowinFormat is as follows:

```
Butanol: CCCCCO
Benzene: c1ccccc1
Fluorobenzene: c1ccccc1F
Acetone: CC(=O)C
```

**Save User**

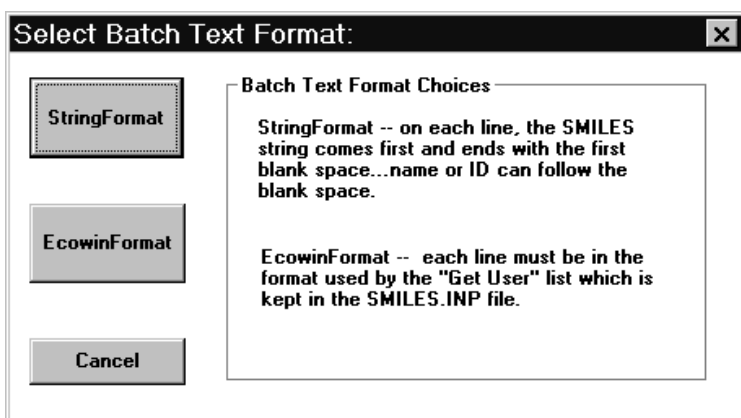
**F6:** Pressing the F6 key or clicking the "Save User" button displays a file selection dialog box that allows the user to save the SMILES notation and chemical name currently showing on the data entry screen to the file. The default name of the file is SMILES.INP; this is for compatibility with similar estimation programs. After a file is selected (or entered by the user), ECOSAR appends the SMILES notation and chemical name currently showing on the data entry screen to the file. If the file does not already exist, ECOSAR will create it and append the current SMILES and name as the first entry. The SMILES and names in a "Saver User" file can be accessed from the data input screen by pressing the F4 key. See Appendix B.

**F7:** The F7 key is used to enter CAS numbers from an ascii text file...the number of CAS numbers in the file is not limited. The user must enter the file name....a election menu is not currently available. The F7 key is used primarily for batch-mode runs...output is written to files named "CASLOG#.OUT" where "#" is a number determined by the program.

The format of the ascii text file is: no spaces in front of the CAS number, hyphens and leading zeros are optional, and a trailing cartridge return....example:

```
000050-00-0
71-43-2
108883
000050-02-2
```

**NOTE: the presence of SRC's SMILECAS.DB database is required!** It is not included with ECOSAR Class Program unless acquired separately.



### CAS Input

**F8:** Pressing the F8 key or clicking the "CAS Input" button requires the presence of a supplemental database file (SMILECAS.DB) and index file in the current subdirectory. A small data entry window is created on the data entry screen which asks for the CAS number of the chemical. An error message will appear in the window if the program can not find the database or index file. The database file contains about 103,000 entries, but not all chemicals with CAS numbers are included in the file. If the chemical is not in the database, an appropriate message is displayed. The program can identify impossible CAS numbers by examining the check digit (the final number of the CAS). The SMILECAS Database is not included with the ECOSAR Class Program installation. It must be acquired and installed separately (Syracuse Research Corp., Environmental Science Center).

### Calculate

**PgDn:** Pressing The PgDn key or clicking the "Calculate" button calculates the SMILES currently showing on the data entry screen. If an acceptable SMILES has been entered, the Results Window will either appear or be updated. If an incorrect SMILES has been entered, an error message box will appear. After removing the error message box, the incorrect SMILES can be recalled and then edited by pressing the F2 key or clicking the "Previous" button.

**Esc:** During data entry, pressing the Esc key exits the program. When the Results Window is active, pressing the Esc key removes the Results Window.

**Enter:** Pressing the Enter (Return) key sends the cursor to the next data entry field.

**Tab or Shift-Tab:** changes entry fields.

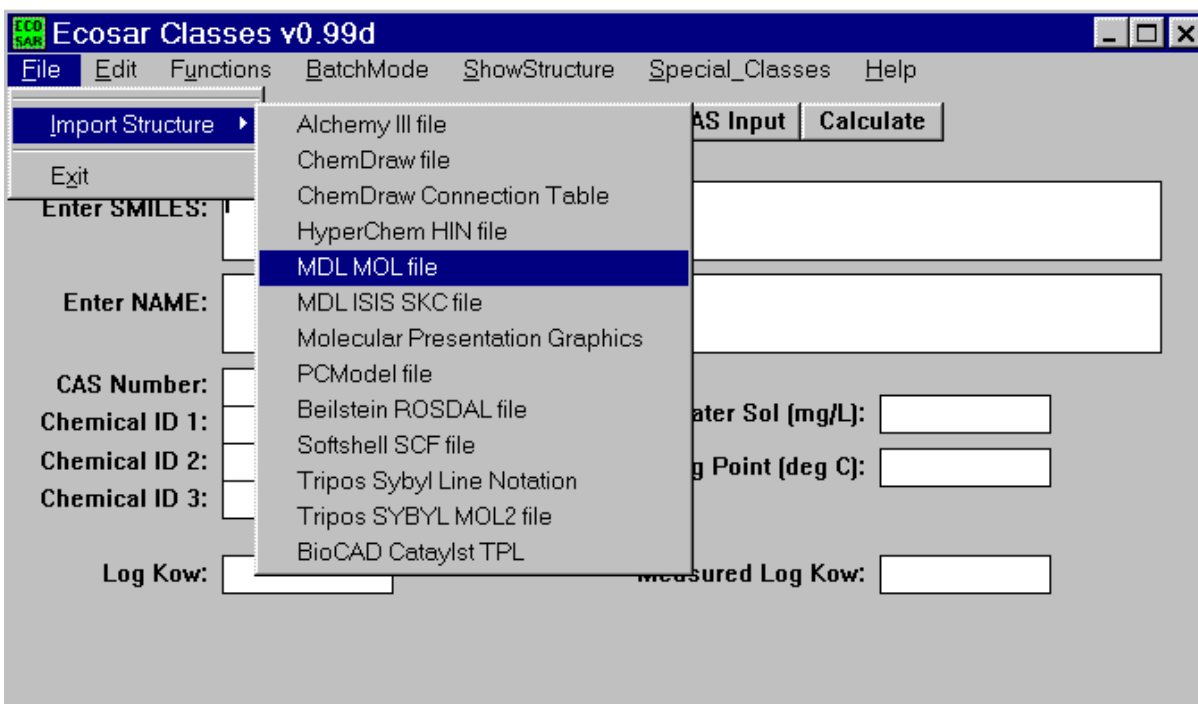
## 5.3. Importing Structures

*Note: this feature is available only when the KOWWIN program is located in the same subdirectory as the ECOSAR program.*

ECOSAR requires a chemical structure in a "SMILES notation" format. ECOSAR (v0.99c and above) adds an "import" features that allows other chemical structure formats to be imported directly into ECOSAR. The "import" feature is accessed from the Menu Bar via: "File"...."Import Structure" as shown in the figure below.

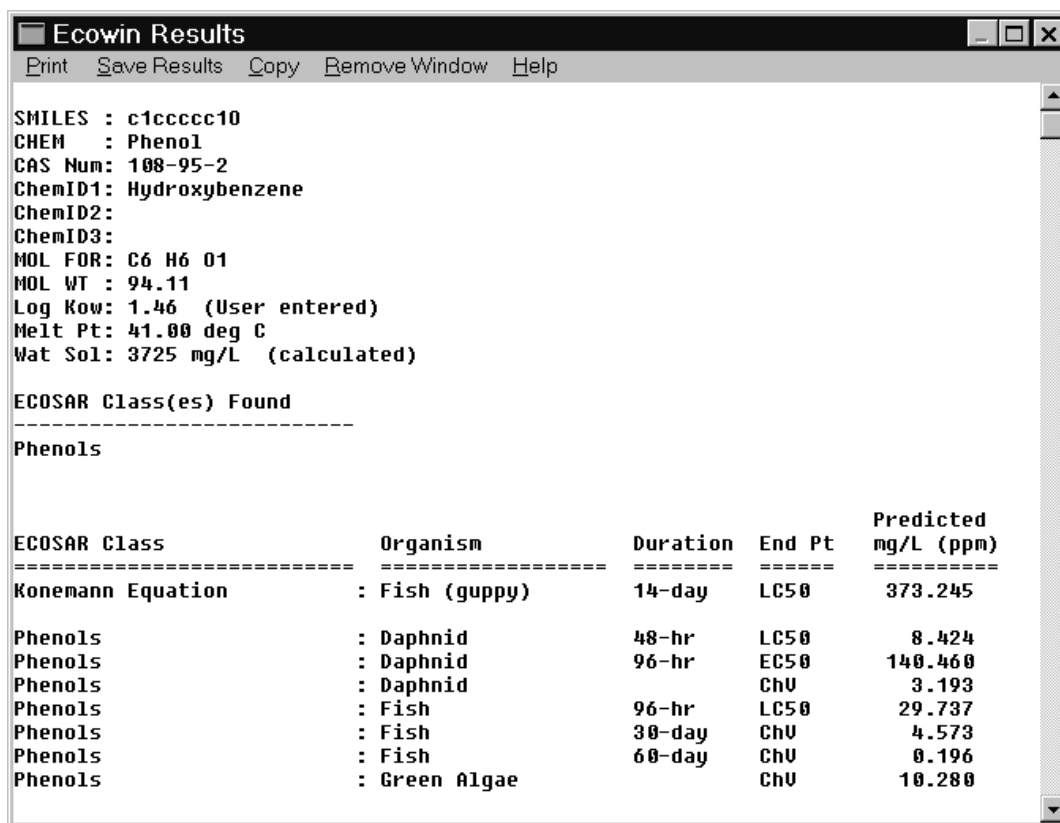
The "import" features uses the structure format conversion engine of the commercial software package ConSystant(tm) available from ExoGraphics, PO Box 655, West Milford, NJ 07480, (201) 728-0188. Syracuse Research Corporation has a license agreement with ExoGraphics that permits incorporation of the ConSystant(tm) DLL with SRC estimation programs. Imported structures are converted to SMILES notations and placed in the SMILES data entry field of ECOSAR. ECOSAR filters the conversion to make the ECOSAR notation as compatible as possible with ECOSAR. However, some converted SMILES notations (especially SMILES with charged ions) will require some user modification before ECOSAR can estimate the structure. Importable structure formats include:

Alchemy III MOL files  
ChemDraw files  
ChemDraw Connection Tables  
HyperChem HIN files  
MDL MOL files  
MDL ISIS SKC files  
Molecular Presentation Graphics MPG files  
PCModel files  
Beilstein ROSDAL files  
Softshell SCF files  
Tripos Sybyl Line Notations  
Tripos SYBYL MOL2 files  
BioCAD Cataylst TPL files



## 6. RESULTS

The Results Window presents the results of ECOSAR Class Program's estimations. It appears when a SMILES notation is calculated. Figure 3 below illustrates an example Results Window:



**Figure 3.** Example Results Screen

The Results Window can be moved, sized and placed anywhere on the Microsoft Windows desktop. It does not need to be removed to calculate another SMILES notation; the Results Window will be updated when another SMILES is calculated.

The Results Window lists the SMILES (which might have been modified by the program due to aromatic detection or other conversion), molecular formula, molecular weight, and the fragments used to derive the estimation. The following menu choices are available when the Results Window is active:

**Print:** prints the results as shown.

**Save Results:** saves the summary output to a file. The output files are named ECOW\*.DAT where "\*" is a number from 1 to 100. Numbering begins at 1 and automatically proceeds to number 100. Currently, all results are appended to the same file number until the program is exited. The next time the program is started, the next available number is used; therefore, different files are used from session to session! If all numbers have been used in existing files, then number 1 will be used and the existing file ECOW001.DAT will be overwritten!!

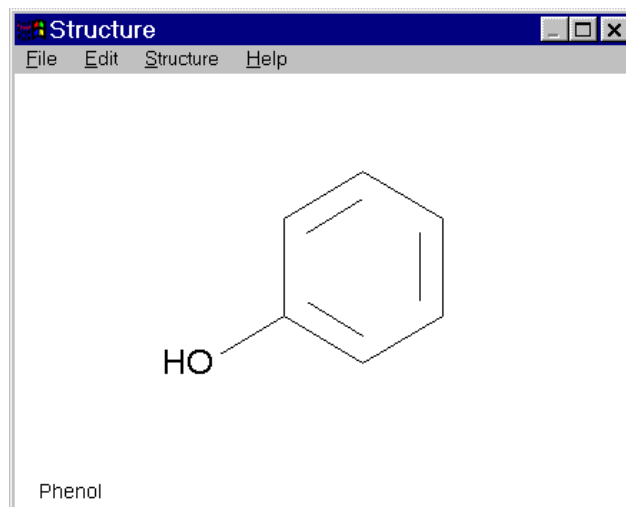
**Copy:** copies the results as shown (minus the rectangle enclosing the estimate) to the Windows clipboard. The results can then be copied into other Windows programs such as word processors. When copied to a word processor (such as Word Perfect, Ami Pro, or Microsoft Word), a non-proportional font (such as courier) must be used for correct formatting!! ... Also, the page width margins must be wide enough!

**Remove Window:** deletes the Results Windows; a new Results Window will appear with the next estimation. It may be more convenient to move and size the Results Window for personal preference (after the first estimation) rather than to remove it after each estimation. If the Results Window is left on the screen, the next estimation results will simply replace the existing results.

The Log Kow value in the Result Window designates whether it was entered by the user or calculated by the KOWWIN program. The Water Solubility value designates whether it was calculated or measured (see Appendix D for water solubility estimation methodology).

### 6.1. Structure Window

*Note: this feature is available only when the KOWWIN program is located in the same subdirectory as the ECOSAR program.* The Structure Window shows a 2-dimensional plot of the chemical structure. An example "Structure" window is shown here. The window shows the entire structure (it does not "clip" sections of the molecule). In order to fit the entire structure in the window, the aspect ratio of the MS-Windows metafile depiction



has been rendered proportional (that is, by changing the height or width of the window, the structure scaling changes). At times, the height or width of the window may need to be changed to give a better structure depiction. When results from the Results window are printed with the "Print results with structure" option, the aspect ratio of the structure will be printed (if possible) with the same aspect ratio as the Structure window.



The Structure Window Menu Bar gives access to printing the structure, saving the structure as an MDL MOL file or an ISIS SKC file, copying the structure to the MS-Windows clipboard, or changing selected window parameters. Changeable windows parameters include background colors of the structure or bottom text areas. Double clicking the text at the bottom of the window allows the text to be changed. Copying the structure (from the menu bar Edit) to the Windows clipboard has two options:

- (1) Copy (as placeable metafile): this copies both structure and text to the clipboard. Some word processors and drawing programs require "placeable metafiles" for graph import. The ability of other Windows programs to use placeable metafiles varies.
- (2) Copy structure (as metafile): the copies only the structure to the clipboard. Most commercial word processors will import this format.

## 6.2. Konemann Equation

The Konemann equation is an equation developed from a variety of different compounds (including chlorobenzenes, chlorotoluenes, chloroalkanes, diethyl ether and acetone) using guppies and 14-day exposure periods (Konemann, 1981). The equation is:

$$\text{Log (1/LC50)} = 0.871 \log K_{ow} - 4.87$$

where LC50 is in  $\mu\text{mol/L}$ .

## 6.3 Example SAR Equations

The following are example SAR equations used by the ECOSAR Class Program to calculate ecotoxicity values. They are indicative of all SARs calculated from SMILES and  $\log K_{ow}$  values.

### Acrylates:

$\text{Log 48-h LC50} = 0.00886 - 0.51136 \log K_{ow}$	(Daphnids, mortality)
$\text{Log 96-h LC50} = -1.46 - 0.18 \log K_{ow}$	(Fish, mortality)
$\text{Log ChV} = -1.99 - 0.526 \log K_{ow}$	(Fish chronic value; survival/growth)
$\text{Log 96-h EC50} = -1.02 - 0.49 \log K_{ow}$	(Green Algae, growth)

The values calculated by these equations are in units of millimoles/L.

## 7. BATCH RUNS

Batch runs are used to make multiple estimates from a single input file. The ECOSAR Class Program can make "batch runs" from three different types of input files. Each input file must be in a specific format, otherwise, the batch run will fail. Program access to "batch-runs" is available from (a) the top menu option "BatchMode", (b) various options under the top menu option "Functions", and (c) the F5, F7 Function keys. The following describes each "batch run" input file that ECOSAR Class Program can use.

**(1) CAS Number List** - This is a plain text file (usually with a ".txt" file extension) containing a list of CAS (Chemical Abstract Service) Registry numbers. The format of the ascii text file is: no spaces in front of the CAS number, hyphens and leading zeros are optional, and a trailing carriage return. For example:

```
000050-00-0
71-43-2
108883
000050-02-2
```

NOTE: the presence of SRC's SMILECAS.DB database is required! The SMILECAS Database must be in the same subdirectory as the ECOSAR Class Program program. There is no limit to the number of CAS numbers in the file. The F7 function key accesses the CAS batch list option.

**(2) SMILES String, String Format List** - This is a plain text file (usually with a ".txt" file extension) containing a list of SMILES notations. A "String Format" list must have the SMILES string at the beginning of each line in the file; it can then be followed by a space(s) and then the name or other ID. The SMILES string is considered terminated at the first space. An example String Format is as follows:

```
CCCCO Butanol
c1ccccc1 Benzene
Fc1ccccc1 Fluorobenzene
CC(=O)C Acetone
```

The F5 function key accesses the SMILES String batch option. The output file is named "BATCH#.OUT" where "#" is a number determined by the program.

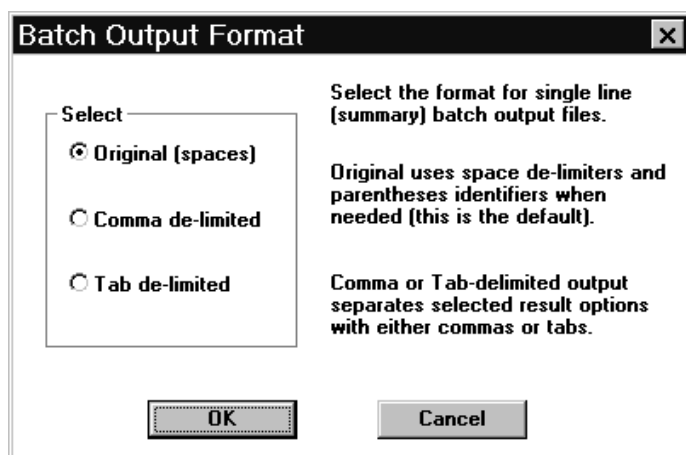
**(3) SMILES String, Ecosar Format List** - This is a plain text file (usually with a ".inp" file extension) containing a list of SMILES notations. EcwinFormat is the same format used by the "Get User" and "Save User" button features. Therefore, the "SMILES.INP" file can be used directly to run batch file outputs. In this format, the name comes first (maximum of 60 characters) followed by a colon and one space, and then the SMILES notation. An example Pckoc Format is as follows:

Butanol: CCCCCO  
Benzene: c1ccccc1  
Fluorobenzene: c1ccccc1F  
Acetone: CC(=O)C

The F5 function key accesses the SMILES String batch option. The output file is named "BATCH#.OUT" where "#" is a number determined by the program.

### 7.1. Batch Output Formats

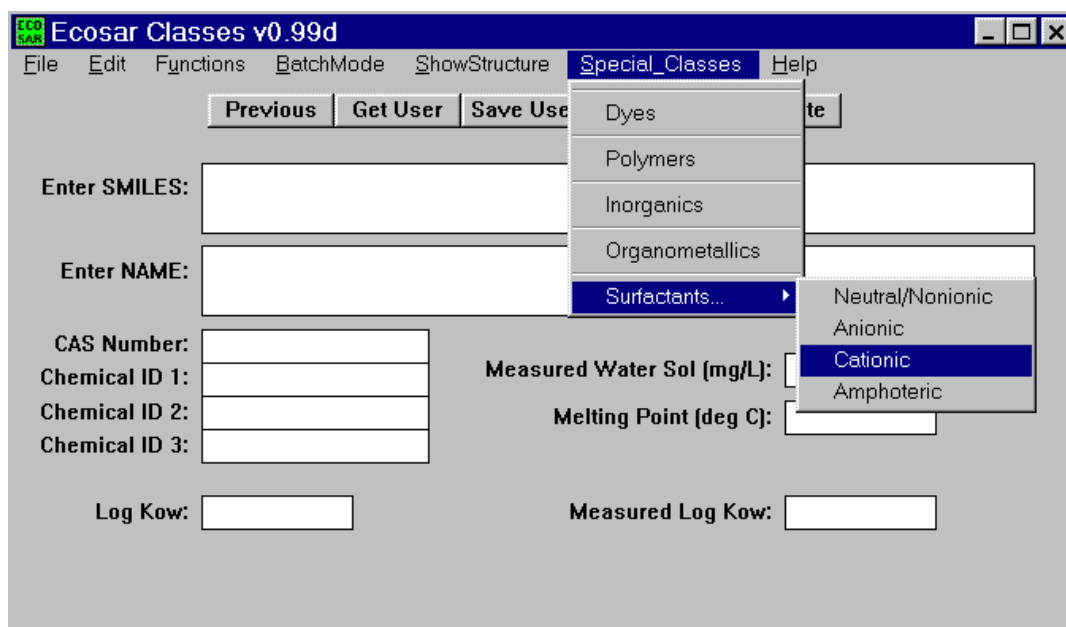
Batch runs can capture results as either "Full Output" or "Summary Output". Full Output captures results for each compound the same as they would appear in the "Result Window" (if each compound was estimated individually); these output files can get very large for large numbers of compounds. Summary Output captures selected results and places these results on a single line for each compound. Before running a batch with "Summary Output", the format of the output file can be selected from the dialog box shown here. The default is "space filled" with required identifiers to identify various results. Output can also be "Comma de-limited" or "Tab de-limited". These output selections separate results on each line with either commas or tabs. This is useful for importing batch output file directly into other programs( such as Microsoft Excel™ or Lotus123™ spreadsheets).



## 8. SPECIAL CLASS CALCULATIONS

The ECOSAR Class Program has been developed primarily for the following scenario:

(1) enter a SMILES notation, (2) computer determination of appropriate ECOSAR classes for the SMILES notation, and (3) calculate the ecotoxicity SARs using a log  $K_{ow}$  value. Several "Special Classes" of ECOSAR SARs or classifications do not use the log  $K_{ow}$  value or can not be adequately classified from the SMILES (in this ECOSAR version). These "Special Classes" include (a) Polymers, (b) Inorganics, (c) Dyes, and (d) Surfactants. The current version of the ECOSAR Class Program does not include SARs for Polymers, Dyes, or Inorganics (these may be added in the future). However, SARs are available for various Anionic, Cationic, Nonionic, and Amphoteric Surfactants. Instead of the log  $K_{ow}$  value, these SARs utilize the number of ethoxylate units or the average length of a carbon chain. These "Special Classes" are accessed from the Main Menu bar (see Figure 4).



**Figure 4.** Special Class Menu Options

The Special Classes have their own data entry dialog box (see Figure 5). The calculated results are placed in the same Results Windows as results using SMILES notations (an example is illustrated in Figure 6). **Note:** the Water Solubility or Water Dispersibility fields in the data entry dialogs are not used in SAR calculations.

**Amphoteric Surfactants**

Chemical Name:

CAS Num:

Chemical ID 1:

Chemical ID 2:

Chemical ID 3:

Water Sol (mg/L):

Num Ethoxylates:

Highlight Amphoteric Surfactant Class.  
Press Calculate Button. (Note:  
SMILES Not Required or Used)

Alkyl-Nitrogen-Ethoxylates  
Surfactants, Ethomeen (C=08)  
Surfactants, Ethomeen (C=09)  
**Surfactants, Ethomeen (C=10)**  
Surfactants, Ethomeen (C=11)  
Surfactants, Ethomeen (C=12)  
Surfactants, Ethomeen (C=13)  
Surfactants, Ethomeen (C=14)  
Surfactants, Ethomeen (C=15)  
Surfactants, Ethomeen (C=16)  
Surfactants, Ethomeen (C=17)  
Surfactants, Ethomeen (C=18)

**Figure 5.** Example Entry Dialog Box for Surfactants

**Ecwin Results**

Print Save Results Copy Remove Window Help

CHEM : Surfactant P100  
CAS Num:  
ChemID1:  
ChemID2:  
ChemID3:  
Wat Sol: 0 mg/L

Number of Ethoxylates: 3.00

ECOSAR Class: Surfactants, Ethomeen (C=10)

Organism	Duration	End Pt	Predicted mg/L (ppm)
=====	=====	=====	=====
Daphnid	48-hr	LC50	7.745
Fish	96-hr	LC50	7.745
Algae	96-hr	EC50	7.745

**Figure 6.** Example Results Window for Surfactants

## 9. BIBLIOGRAPHY

- Koneman, H. 1981. Fish toxicity tests with mixtures of more than two chemicals: a proposal for a quantitative approach and experimental results. *Toxicology* 19: 229-238.
- Meylan, W.M. and P.H. Howard. 1994a. Upgrade of PCGEMS Water Solubility Estimation Method (May 1994 Draft). prepared for Robert S. Boethling, U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics, Washington, DC; prepared by Syracuse Research Corporation, Environmental Science Center, Syracuse, NY 13210.
- Meylan, W.M. and P.H. Howard. 1994b. Validation of Water Solubility Estimation Methods Using Log  $K_{ow}$  for Application in PCGEMS & EPI (Sept 1994, Final Report). prepared for Robert S. Boethling, U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics, Washington, DC; prepared by Syracuse Research Corporation, Environmental Science Center, Syracuse, NY 13210.
- Meylan, W.M. and Howard, P.H. 1995. Atom/Fragment contribution method for estimating octanol-water partition coefficients. *J. Pharm. Sci.* 84: 83-92.
- Meylan, W.M. and Howard, P.H. 1996. Improved method for estimating water solubility from octanol/water partition coefficient. *Environ. Toxicol. Chem.* 15: 100-106.
- Weininger, D. 1988. SMILES, A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **28**: 31-36.

## APPENDIX A

### Selected SMILES Information

A SMILES notation is considered terminated at the first blank space. Characters following the first blank space are ignored!

#### Entering the Nitro Function (NO<sub>2</sub>)

The nitro function (NO<sub>2</sub>) is usually written as N(=O)(=O) or N(=O)=O in SMILES notation. In this program version, the nitro function can also be designated (simply) by the capital letter **T**.

#### Entering the Sulfonic Acid Function

The sulfonic acid function (-SO<sub>2</sub>-OH) is usually written as S(=O)(=O)O in SMILES notation.

#### Carbonyl Function (C=O) Information

The carbonyl function (C=O) should always be entered in upper case letters. Additional information is presented in the SRC document "A Brief Description of SMILES Notation".

#### Metals & Charged Species

Charged species can not be entered directly into the program with + and signs. Compounds, such as QACs (quaternary ammonium compounds), can be entered by simply attaching the charges as if a direct bond exists; for example, tetramethyl ammonium bromide can be entered as > N(C)(C)(C)(C)Br ...also, for many hydrochlorides, simply ignore the HCL portion of structure (leave it out and enter the compound as the nonhydrochloride; alternatively, see section below).

ECOSAR Class Program can accept and evaluate the following METALS:

Na sodium    Hg mercury    K potassium    Li lithium

Use the chemical symbol to include any of these metals; for example, sodium acetate could be: NaOC(=O)C. Alternatively, the above metals and ALL OTHER metals can be put in a SMILES notation by bracketing as follows: [Na] sodium    [As] arsenic    [Ca] calcium    [Sn] tin    [Pb] lead .... etc.

Valence charges are NOT evaluated in brackets!! ATTACH metals to the corresponding negatively charged species and do NOT use + and charges in the SMILES!! Example: in some SMILES notations, sodium hexanoate would be entered as: [Na+][O]C(=O)CCCCC however, this is not allowed in this program because charges are not allowed and oxygen can not be bracketed.

#### Entering Hydrogen Directly

For ECOSAR Class Program, direct hydrogen entry in a SMILES notation is not allowed with the exception of connection to aliphatic or aromatic nitrogen for the purpose of entering a nitrogen with a valence greater than +3 (eg, various quaternary ammonium compounds and hydrochlorides)...nitrogens with a valence of +3 or less ignore direct hydrogen entries. Hydrogen is entered as an upper case H (as in the following examples):

(1) acridine hydrochloride: c1ccc2cc3ccccc3n(H)(CL)c2c1

(2) benzenepentanamine hydrochloride: c1ccccc1CCCCN(H)(H)(H)CL

When to include the "HCL" in SMILES for various hydrochlorides depends upon the nature of the hydrochloride...for example, most hydrochlorides represented generically as: Formula HCL can ignore the HCL; however, most ammonium-type compounds (such as #2 above) require the direct hydrogens.

#### Aromatic Selenium

Aromatic selenium can be entered as either (1) lower case se or (2) as [se] ....for example, selenofuran could be entered as (1) c1csecc1 or as (2) c1c[se]cc1 ....if entered as: C1=CSeC=C1, ECOSAR Class Program will automatically convert it to: c1c[se]cc1

#### Miscellaneous

In selected diazoacetyl compounds (eg, azaserine, N2=CH-CO-O-CH2-CH(-NH2)-COOH), the N<sub>2</sub> is commonly written as: N+=N-. For the purposes of SMILES notation, the unit is considered as: N#N.

## **APPENDIX B**

### **Description of the User Input File**

The User Input File is a file containing up to 1500 SMILES notation and chemical names that can be accessed during the execution of ECOSAR Class Program. It can be used to enter SMILES notations and chemical names onto the data entry screen. By default, the User Input File is named SMILES.INP. This name must be used; it can not be changed by the user. The 1500 entries that comprise SMILES.INP are determined by the user. This file can be useful for purposes other than data entry into ECOSAR Class Program. For example, it can be used for record keeping purposes. It can also be used for entering data into other estimation programs available from Syracuse Research Corporation that utilize SMILES.INP and SMILES notation, such as HENRYWIN (estimation of Henry's Law Constant), AOPWIN (estimation of atmospheric oxidation) and KOWWIN (estimation of octanol-water partition coefficient).

The User Input File is accessed during ECOSAR Class Program data entry by pressing the F4 key. The SMILES.INP file must exist in the subdirectory from which ECOSAR Class Program was started.

The SMILES notation and chemical name showing on the data input screen can be added to the SMILES.INP file by pressing the F6 key during data entry. If the SMILES.INP file doesn't already, the F6 key will create it and add the current notation and name as the first entry. Currently, there is no way to edit or delete entries to SMILES.INP during ECOSAR Class Program. However, SMILES.INP is a plain text file and it can be edited with any text editor or word processing program (as long as it is imported and saved as a DOS text file). Any text editor or word processing program can be used to create and add entries to SMILES.INP as long as the format is correct. The correct format is the following: the chemical name (up to 60 characters) followed by a colon (:), then one space (and only one space) followed by the SMILES notation and a carriage return.

## **APPENDIX C**

### **CAS Number Data Base**

The CAS Number data base is used to input SMILES notations and chemical names onto the data entry screen by entering the Chemical Abstract Service (CAS) Registry number of a chemical. It is available as a separate product from Syracuse Research Corporation and is not included with the ECOSAR Class Program. The CAS Number data base (SMILECAS.DB) and index file (SMILECAS.IDX) must be located in the subdirectory from which ECOSAR Class Program was started. A hard disk is required to use the data base due to the size of the data base. The SMILECAS.DB file is approximately 7.3 MB and the index file is approximately 2.4 MB.

The CAS Number data base currently contains 103,000 entries. The initial 20,000 entries were obtained from the U.S. EPA file of CAS numbers, SMILES notations and chemical names used by the GEMS program software. The entries in this file are the discrete organics listed in the U.S. EPA TSCA Inventory. Although the number of entries is large, various chemicals that may be of interest may not be included in the data base.

The CAS Number data base is accessed by pressing the F8 key at the data entry screen. A pop-up window will appear requesting entry of the CAS number.

The SMILECAS.DB file is a translated version of a dBase® III+ DBF file. The dBase® DBF file is not used by ECOSAR Class Program due to the inefficient space filling which exists in a DBF file (the DBF file is about 35 MB in size compared to 7.3 MB for the DB file).



## APPENDIX D

### Estimation of Water Solubility

The ECOSAR Class Program estimates water solubility using methodology developed for the U.S. EPA and described in Meylan and Howard (1994a, 1994b, 1996). The estimation equations used in the current version are as follows:

No Melting Point Available:

$$\log \text{WaterSol (moles/L)} = -0.312 - 1.02 \log K_{ow}$$

Liquid at 25 deg C:

$$\log \text{WaterSol (moles/L)} = 0.551 - 1.091 \log K_{ow}$$

Solid at 25 deg C:

$$\log \text{WaterSol (moles/L)} = 0.2236 - 1.009 \log K_{ow} - 0.00956 (T_m - 25)$$

(where  $T_m$  is the melting point in deg C)

**Note:** all water solubility estimates pertain to 25 deg C.

## APPENDIX E

### List of ECOSAR Chemical Classes

The following is an alphabetic list of chemical classes identified from SMILES notations by the ECOSAR Class Program:

Acid Chloride/Halide	Neutral Organics
Acrylamides	Peroxy Acids
Acrylates	Phenols
Aldehydes	Phenols (dinitro)
Aliphatic Amines	Propargyl Alcohols
Anilines (amino-meta)	Propargyl Alcohols - Hindered
Anilines (amino-ortho)	Propargyl Ethers
Anilines (amino-para)	Quinone
Aromatic Amines	Salicylates
Azides	Salicylic Acid
Aziridines	Schiff Bases
Benzotriazoles	Silamines
Benzyl Alcohols	Silanes (alkoxy)
Benzyl Amines	Surfactants-anionic
Benzyl Halides	Surfactants-cationic
Diazoniums	Surfactants-nonionic
Diepoxides	Thiazolidinones
Diketones	Thiazolinone (iso-)
Dinitro Aromatic Amine	Thiocyanates
Dinitrobenzenes	Thiols(mercaptans)
Epoxides	Thiophenes
Esters	Triazines
Esters (phosphate)	Ureas(substituted)
Haloacetamides	Vinyl/Allyl Alcohols
Hydrazines	Vinyl/Allyl Ethers
Imides	Vinyl/Allyl Halides
Isocyanates	Vinyl/Allyl Ketones
Malononitriles	Vinyl/Allyl Sulfones
Methacrylates	